

CLAIMS

1. A method for extracting visemes from a speech signal, comprising:
receiving successive frames of digitized analog speech information obtained from
5 the speech signal at a fixed rate;
filtering each of the successive frames of digitized analog speech information to
synchronously generate time domain frame classification vectors at the fixed rate,
wherein each of the time domain frame classification vectors is derived from one of the
successive frames of digitized analog speech information; and
10 analyzing each of the time domain classification vectors to synchronously
generate a set of visemes corresponding to each of the successive frames of digitized
speech information at the fixed rate.
2. The method for extracting visemes from a speech signal according to claim 1, wherein
15 in the step of analyzing, each set of visemes is generated with a latency less than 100
milliseconds with reference to a successive frame of digitized analog speech information
with which the set of visemes corresponds.
3. The method for extracting visemes from a speech signal according to claim 2, wherein
20 the latency is less than 10 milliseconds.
4. The method for extracting visemes from a speech signal according to claim 1, wherein
each set of visemes includes a subset of visemes identifiers and a one to one
corresponding subset of confidence numbers.
25
5. The method for extracting visemes from a speech signal according to claim 1, wherein
the set of visemes consists of an identity of one most likely viseme.
6. The method for extracting visemes from a speech signal according to claim 1, wherein
30 the step of filtering comprises:
converting each of the successive frames of digitized analog speech information
to a spectral domain vector using N multi-taper discrete prolate spheroid sequence basis
(MTDPSSB) functions that are factors of a Fredholm integral of the first kind; and
converting each spectral domain vector to one of the time domain frame
35 classification vectors using Inverse Discrete Cosine Transformation, wherein N is a
positive integer.

7. The method for extracting visemes from a speech signal according to claim 6, wherein the conversion of each of the successive frames of digitized analog speech information to a spectral domain vector comprises:

5 multiplying a successive frame of digitized analog speech information by one of the N MTDPSSB functions to generate N product sets of the successive frame of digitized analog speech information;

 performing a fast Fourier transform (FFT) of each of the N product sets to generate N FFT sets of the successive frame of digitized analog speech information; and

10 adding (change adding to combining because the addition is done to magnitude spectrums rather than separately to the real and imaginary components) together the N FFT sets of the successive frame of digitized analog speech information to generate a summed FFT set of the successive frame of digitized analog speech information.

15 8. The method for extracting visemes from a speech signal according to claim 1, wherein the conversion of each of the successive frames of digitized analog speech information to a spectral domain vector further comprises scaling the summed FFT set of the successive frame of digitized analog speech information.

20 9. The method for extracting visemes from a speech signal according to claim 1, wherein the step of analyzing comprises a spatial classification.

25 10. The method for extracting visemes from a speech signal according to claim 1, wherein the step of analyzing is performed by one of a neural network and a fuzzy logic function.

30 11. The method for extracting visemes from a speech signal according to claim 9, wherein the neural network is a feed-forward memory-less perceptron type neural classifier.

12. An apparatus for extracting visemes from a speech signal, comprising:

 at least one processor; and

 at least one memory that stores programmed instructions that control the at least one processor to

35 receive successive frames of digitized analog speech information from the speech signal at a fixed rate,

filter each of the successive frames of digitized analog speech information to synchronously generate time domain frame classification vectors at the fixed rate, wherein each of the time domain frame classification vectors is derived from one of the successive frames of digitized analog speech information, and

5 analyze each of the time domain classification vectors to synchronously generate a set of visemes corresponding to each of the successive frames of digitized speech information at the fixed rate.

10 13. A speech receiving device, comprising:

 at least one processor;

 at least one memory that stores programmed instructions that control the at least one processor to

 receive successive frames of digitized analog speech information from a
15 speech signal at a fixed rate,

 filter each of the successive frames of digitized analog speech information to synchronously generate time domain frame classification vectors at the fixed rate, wherein each of the time domain frame classification vectors is derived from one of the successive frames of digitized analog speech information, and

20 analyze each of the time domain classification vectors to synchronously generate a set of visemes corresponding to each of the successive frames of digitized speech information at the fixed rate; and

 a display that displays an avatar that is formed using the set of visemes.

25 14. An apparatus for extracting visemes from a speech signal, comprising:

 means for receiving successive frames of digitized analog speech information from the speech signal at a fixed rate,

 means for filtering each of the successive frames of digitized analog speech information to synchronously generate time domain frame classification vectors at the
30 fixed rate, wherein each of the time domain frame classification vectors is derived from one of the successive frames of digitized analog speech information, and

 means for analyzing each of the time domain classification vectors to synchronously generate a set of visemes corresponding to each of the successive frames of digitized speech information at the fixed rate.